John von Neumann Institute for Computing

Binding Properties in Protein Nucleic Acids

B. Dietrich, T. A. Hupp, B. Engels

published in

NIC Symposium 2001, Proceedings,
Horst Rollnik, Dietrich Wolf (Editors),
John von Neumann Institute for Computing, Jülich,
NIC Series, Vol. **9**, ISBN 3-00-009055-X, pp. 63-72, 2002.

http://www.fz-juelich.de/nic-series/volume9

Binding Properties in Protein Nucleic Acids

B. Dietrich, T. A. Hupp, and B. Engels

Institut für Organische Chemie, Universität Würzburg
Am Hubland, 97074 Würzburg, Germany

The genetic information of life is encoded in the sequence of the bases of the DNA. In B-DNA, the dominant conformer of the DNA under physiological conditions, the base topology, i.e. the orientation and available space for the bases, is strictly limited by the helical structure, in which both strands wind around a common axis in opposite directions (Figure 1).

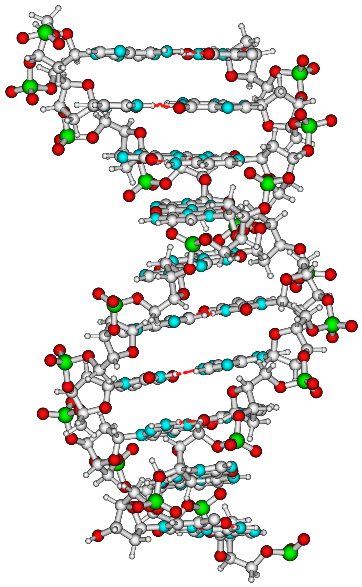


Figure 1. Helical double strand in B-DNA

As a consequence, in B-DNA a purine-base always pairs with a pyrimidine-base. While guanine always pairs with cytosine adenine always pairs with thymine. Furthermore all pairs use the Watson-Crick-pairing-mode, as shown in Figure 2 and 3.

The structure and stability of such systems are determined by a variety of different interaction[1]. Since the helical structure of DNA has been clarified by Watson and Crick, hydrogen bonds have been considered to represent the determinant factor for the stability of base paired systems and for the selectivity of the pairing. In the gas-phase, it has been indeed found experimentally and by calculations, that each hydrogen bond has a strength of 5-7 kcal/mol[2,3]. In aqueous medium, however, a hydrogen bond contributes only 0.5-1.8 kcal/mol[4–6]. The weakness of hydrogen bonds in water results because the

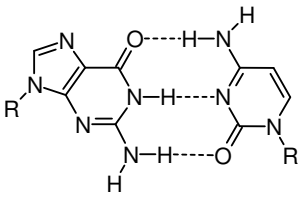


Figure 2. Guanine-cytosine-pair in DNA

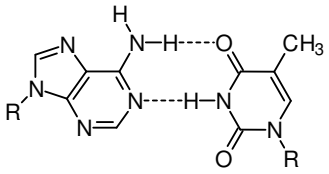


Figure 3. Adenine-thymine-pair in DNA

nucleic bases form strong bonds to themselves and to water molecules. Van-der-Waals and dipole interaction between the planar base pairs lead to the so-called π- or base stacking interaction which is the second decisive interaction enabling the formation of duplexes[7]. Van-der-Waals interaction, that is caused by fluctuating dipols, seems to play a central role for this stabilizing factor. In addition, permanent electrostatic attractions and hydrophobic effects contribute to the stacking energy[8–14].

While the hydrogen bonds are formed perpendicular to a single strand between both strands the stacking interaction is mainly oriented along a single strand and consequently stabilises the orientation of a single strand itself. While the strength of hydrogen bonds is strongly weakened in polar solvent the size of the stacking interaction is less influenced. Consequently a correct description of the strength of the base pairing and of the relative importance of the various contributions must include solvent effects. The weakness of the single interaction and the complicated interplay between the various interaction causes the difficulties in the theoretical description of such systems.

To differentiate between the various interaction governing the base pairing process, it is important to know how the geometrical structure of the backbone influences the base pairing itself and the secondary structure of DNA depicted in Figure 4. To obtain information about this topic, one can change the properties of the backbone by replacing functional groups of the sugar-rings. Alternatively one can completely replace the backbone as indicated in Figure 5.

The latter idea does not only open the way to a better understanding of base pairing, but can also be used to create new drugs, if these new DNA analogue substances form more stable double strands with DNA or RNA than with itself. If they are able to bind complementarily to DNA or RNA, this behaviour can be employed to bind to mutated sequences and consequently block the expression of defect genes. In the context of this so-called antisense-strategy, in 1991 a new biopolymer was described[15–18]. In these compounds the complete DNA-sugar-phosphate-backbone is replaced by a peptide-like system consisting of diaminoethylglycine-units (Figure 6).

The diaminoethylglycine-PNA (*p*eptide *n*ucleic *a*cid) pairs well with RNA and is easy to synthesize. The pharmacological application has so far been hampered by the fact that no mechanism exists that transports alanyl-PNA into the nucleus[18,19]. Even though the pharmacological usability is limited so far, PNA's are excellent substances to study base

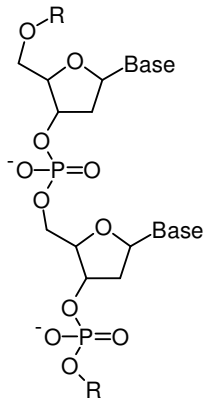


Figure 4. The structural composition of a DNA-strand

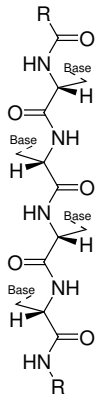


Figure 5. The structural composition of an alanyl-PNA strand

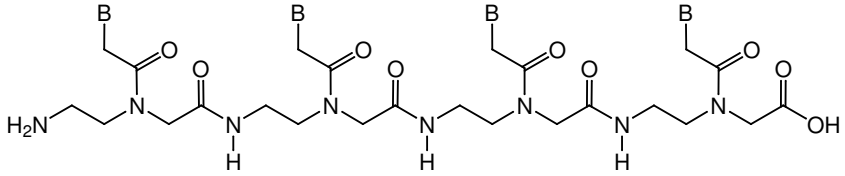


Figure 6. Diaminoethylglycine-PNA (B = base)

pairing, since backbone properties can be varied easily by using different amino acids. Even the secondary structure can be varied strongly. Examples are PNA's being formed from alternating *D*- and *L*-amino acids[20]. They build up double strands which are no longer helical, but possess linear geometrical structure (Figure 7). The distance between two neighbouring nucleic acids within the same strand is 3.6 Å, a value very similar to the favourized distance between stacked bases in the DNA (3.4 Å). Because of the rigidity of the peptidic backbone, the orientation of the nucleic acids is well defined so that the complicated geometrical conditions in the DNA are strongly simplified.

Consequently the alanyl-PNA (Figure 5 and 7) offers an excellent opportunity to differentiate between the three remaining interaction stabilizing a PNA-hexameric duplex, namely the hydrogen bonds between the strands, the van-der-Waals-interaction between neighbouring bases (stacking interaction) and the influence of the surrounding water molecules. In difference to B-DNA in alanyl-PNA systems purines also form base pairs with purines and besides the Watson-Crick pairing mode also the *reverse* Watson-Crick (Figure 8) and the Hoogsten pairing mode can be observed[20] (Figure 9).

In the present project we want to obtain insight into the relative importance of the various interaction governing the double strand forming process of PNA. To achieve this goal we performed all sorts of computations starting from high level quantum chemical

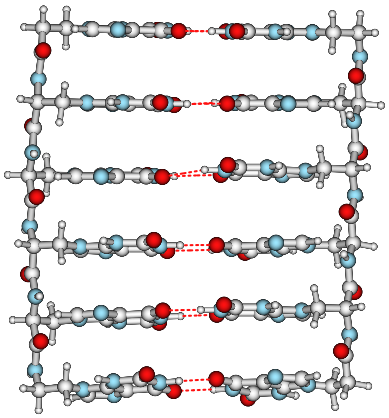


Figure 7. Linear alanyl-xanthine-PNA hexameric duplex

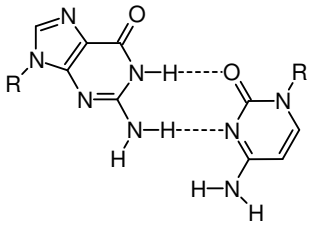


Figure 8. *Reverse*-Watson-Crick base pairing of guanine and cytosine

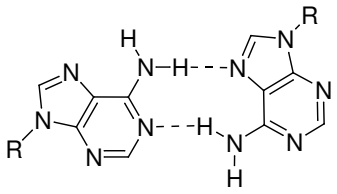


Figure 9. Hoogsten base pairing of adenine und adenine

methods to low level force field ansätze. While the former are made to study the accuracy of the latter, in the force field computations the formation for a whole PNA hexamer in solution can be described. To obtain accurate descriptions high level methods will be used to fit new parameters for the force field. This is necessary, since very accurate parameters are only known for so-called canonical bases. They fail in the description of non-canonical basis such as xanthine (Figure 10).

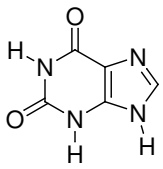


Figure 10. Xanthine

In the present stage of the project we concentrate on xanthine-xanthine alanyl-PNA hexamers which were found to build an unexpectedly stable double strand. The stability of double strands is normally measured by the melting temperature. At the melting temperature the strands form a thermodynamical equilibrium consisting of 50% dissociated and 50% non-dissociated double strands. Changes in the concentration of dissociated vs. non-dissociated double strands can be monitored due to the strong difference between the absorption spectra of oriented double strands and single strands. Alanyl-PNA of nucleic bases which form three hydrogen-bonds normally possess a melting temperature of about 58°C while those bound by only two hydrogen bonds dissociate between 24°C and 32°C. The melting temperature of xanthine alanyl-PNA lies at 48°C, indicating a pairing mode with three hydrogen bonds. However, the normal diketo form of the xanthine molecule can only form two hydrogen bonds to another xanthine molecule opening the question to the reason of this unexpected high binding. To explain the high stability, it was proposed that tautomeric forms (Figure 11) participate in the base pairing. This enables the system to form pairing modes with three instead of two hydrogen bonds (Figure 12). The existence of this kind of pairing mode requires, that the energy of the third hydrogen bond overcompensates the tautomerization energy.

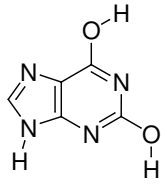


Figure 11. Dienol-xanthine

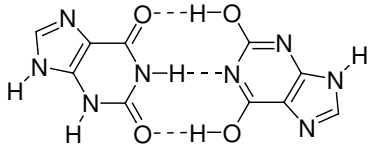


Figure 12. Xanthine-xanthine-base pairing involving dienol-xanthine

The exact pairing geometry cannot be determined by experimental methods, because such substances are synthesized only in very small amounts. Furthermore they are strongly diluted in aqueous solution. Consequently model calculation are important to investigate the base pairing mode of the xanthine-xanthine alanyl-PNA hexamer.

To investigate the role of tautomeric forms, in addition to the interaction discussed for normal double strand forming processes, the energy of the tautomerization has to be considered in the calculations. To see how well the different methods describe the tautomerization energy, we first performed calculations on uracil tautomers (Figure 13, 14 and 15).

Table 1 contains the relative energies of selected tautomeric forms of uracil with respect to the most stable diketo form as a function of various theoretical approaches. It shows that the energy gap between 2-enol-4-keto-3H-uracil and the most stable conformer of the diketo form is about 11-13 kcal/mol. Table 1 also shows that, in comparison to the most accurate CCSD(T) approach, the density functional theory (B3LYP functional) overestimates the energy gap between the various tautomeric forms. One important example

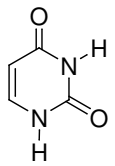


Figure 13. Uracil

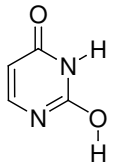


Figure 14. 2-Enol-4-keto-3H-uracil

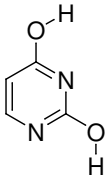


Figure 15. Dienol-uracil

is the dienol form. The energy gaps computed with the more elaborate MP2 approach are very similar to those obtained with the CCSD(T) approach.

Tautomer	Basis set	B3LYP	MP2	CCSD(T)
2-Enol-4-keto-3H	G-31G(d)	12.4	13.1	12.4
	G-31++G(d,p)	11.1	10.5	10.2
Dienol	G-31G(d)	15.3	14.4	14.6
	G-31++G(d,p)	12.8	10.9	10.8

Table 1. Computed relative energies of uracil tautomers. All energies are given with respect to diketo form. (All values are given in kcal/mol.)

Figure 16 shows the six possible twodentate base pairing modes for self pairing of diketo xanthine. The respective relative energies and dimerization energies of all base pairs are given in Table 2. The calculated base pairs show remarkable differences regarding the strength of hydrogen bonds. While the dimerization energies of the *reverse* Watson-Crick and Watson-Crick base pairs **1**, **2** and **3** (E_{dim} = -10.8, -10.4 and -10.6 kcal/mol, respectively) lie in the lower range of the stabilization energies known for twodentate base pairs, the dimerization energy of the strongest bonded base pair N3-*H*/O2 - N3-*H*/O2 (**6**) is nearly twice as high (E_{dim} = -22 kcal/mol). In this base pair both hydrogen bonds involve the N3-*H* group. In the base pairs **4** and **5** which reveal also quite remarkable pairing energies (E_{dim} = -17.1 and -14.1 kcal/mol) only one N3-*H* bond participates in hydrogen bonding while the second donor is provided by the N1-*H* functionality. A comparison of the various base pairing modes shows that the N3-*H* group is essential for the strong hydrogen bond (ca. 5-6 kcal/mol). The nature of the carbonyl groups seem to be less important. As shown by the NBO analyses the unexpected strong hydrogen bonds result from an improved electron transfer from the oxygen lone pairs to the σ*-orbital of the N3-*H*-bond. This can take place because the orbital energy of the σ*-orbital (0.412 a.u.) of the N3-*H* bond is much lower than the orbital energy of the σ*-orbital of the N1-*H* bond (0.442 a.u.). The computed interaction energies between the oxygen lone pairs and the σ*-orbitals of both N-*H* bonds also nicely reflect the variations in the dimerization energies. We obtain about 19 kcal/mol if the N3-*H* bond is involved but only about 12 kcal/mol for the corresponding N1-*H* values. A similar explanation was previously discussed for the strength of the hydrogen bonds in ice.

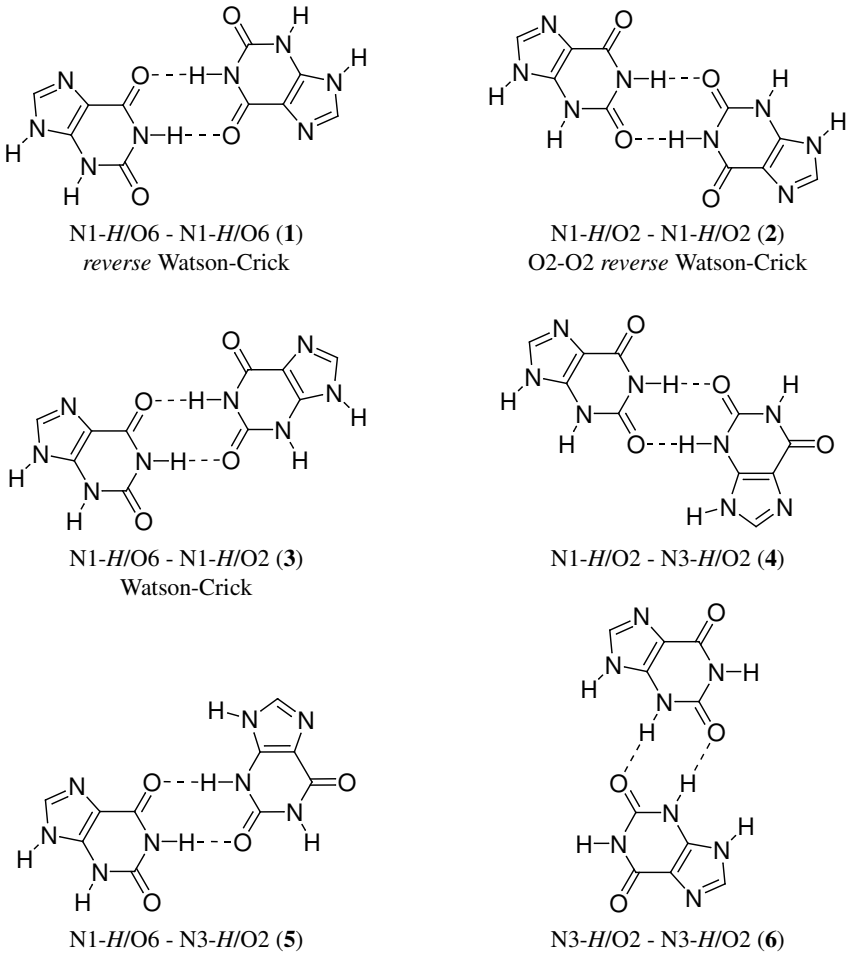


Figure 16. All possibilities for twodentate xanthine-xanthine base pairs with the most stable diketo-xanthine tautomer.

Geometrical constraints on the bonding angles were discussed as possible reasons for variations in the strength of the hydrogen bonds in uracil dimers. Actually, also for xanthine the hydrogen bonds of the more stable pairing mode **6** possesses more idealized bond angles (178.5° for the N-H-O and 121.3° for the H-O-C hydrogen bond) than the less stable hydrogen bonds found in base pairs **1**-**3** (168.7 - 170.8° for the N-H-O and 132.2 to 134.2° for the H-C-O-hydrogen bond). However, if the bond angles of pairing mode **6** are distorted to those optimized for base pairs **1**-**3** the energy raises by only about 2 kcal/mol showing that geometrical constraints are not too much important. To investigate whether tautomeric forms can explain the unexpected high melting temperature of xanthine-xanthine pairing in alanyl-PNA hexamers next to the pairing modes of diketo-xanthine various pairs involving the tautomers 1-H-2-enol-6-keto-xanthine, 3-H-6-enol-2-

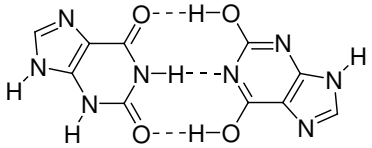
Xanthine-xanthine-base pair	E_{dim}	ΔE_{dim}
N1-*H*/O6 - N1-*H*/O6 *reverse* Watson-Crick **1**	-10.8	0.0
N1-*H*/O2 - N1-*H*/O2 *reverse* Watson-Crick **2**	-10.6	0.2
N1-*H*/O6 - N1-*H*/O2 Watson-Crick **3**	-10.4	-0.4
N1-*H*/O2 - N3-*H*/O2 **4**	-14.1	-3.3
N1-*H*/O6 - N3-*H*/O2 **5**	-17.1	-6.3
N3-*H*/O2 - N3-*H*/O2 **6**	-22.0	-11.2

Table 2. Comparison of MP2/TZVPP-energies of xanthine-xanthine-base pairings. (All values in kcal/mol.)

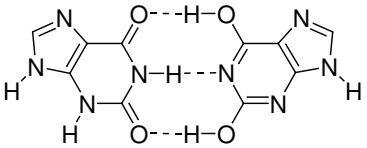
keto-xanthine, and dienol-xanthine were investigated (Figure 17). Base pairs involving other tautomeric forms are energetically too unfavourable. As summarized in Table 3 the tridentate base pairing of diketo-xanthine with dienol-xanthine in the *reverse* Watson-Crick **7** and the Watson-Crick-pairing mode **8** have dimerization energies of -17.4 and -17.5 kcal/mol, respectively. The higer dimerization energies of **7** and **8** simply result from the additional hydrogen bond since the strength of the single hydrogen bonds of base pairs **7** and **8** are similar to those found in the twodentate base pairs **1**-**3**. Since the tautomerization energy necessary to build up the appropriate isomer of the dienol form (4.6 kcal/mol) is smaller than the stabilization arising from the additional hydrogen bond (5-6 kcal/mol) both base pairs **7** and **8** were indeed predicted to be about 2 kcal/mol lower in energy than the twodentate diketo-xanthine complexes **1**-**3**, but still higher than the base pairing modes **4**-**6**. Base pairs that need two xanthine nucleobases in an enol tautomeric form are energetically unfavourable since the overall tautomerization energy is too large. The most stable examples are the base pairs with the 1-H-2-enol-6-keto-xanthine recognizing 3-H-2-enol-6-keto-xanthine (**9**) or 3-H-6-enol-2-keto-xanthine (**10**), respectively. They are predicted to be energetically disfavoured by 1.4 and 4.4 kcal/mol compared to the twodentate base pair **1** although they possess very high dimerization energies of -30.6 and -34.4 kcal/mol.

Xanthine-xanthine-base pair	E_{dim}	ΔE_{dim}
Diketo-dienol *reverse* Watson-Crick **7**	-17.4	-1.9
Diketo-dienol Watson-Crick **8**	-17.5	-2.0
Ketoenol *reverse* Watson-Crick **9**	-30.6	4.4
Ketoenol Watson-Crick **10**	-34.4	1.4

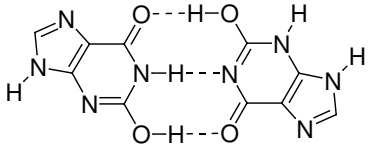
Table 3. Comparison of MP2/TZVPP-energies of xanthine-xanthine-base pairings. (All values in kcal/mol.)



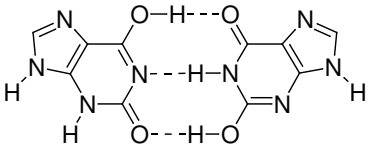
Diketo-dienol *reverse* Watson-Crick (**7**)



Diketo-dienol Watson-Crick (**8**)



Ketoenol *reverse* Watson-Crick (**9**)



Ketoenol Watson-Crick (**10**)

Figure 17. All possibilities for three dentate xanthine-xanthine base pairs with the most stable tautomers of diketoxanthine.

For the base pairs our computations predict the energetic order of pairing modes $\mathbf{6} < \mathbf{5} < \mathbf{4} < \mathbf{8} \approx \mathbf{7} < \mathbf{3} \approx \mathbf{1} \approx \mathbf{2} < \mathbf{10} < \mathbf{9}$. Besides the tridentate base pairs **7** and **8**, also the twodentate pairs **4**-**6** offer explanations for high T_m values found for the xanthine-xanthine double strands. The unusually high T_m values cannot be explained by the twodentate base pairs **1**-**3** since the strength of their hydrogen bonds were found to be in the normal range of twodentate pairing modes.

In the first step of the present project we characterized the hydrogen bonding which represents one of major interaction governing the stability of the double strands formed by xanthine-xanthine alanyl PNA hexamers. We computed all possible xanthine-xanthine pairing modes involving the diketo tautomers and to investigate the importance of enol tautomers all tridentate xanthine-xanthine base pairs lying within the energetic range of the twodentate diketo tautomeric base pairs were also computed. All computations were performed on the MP2/TZVPP//B3LYP/6-31G++G** level of sophistication which was found to be accurate enough for a reliable prediction of the relative energies of the base pairs with and without tautomeric forms[21]. While our model describes the hydrogen bonding between the bases with high accuracy it does not account for other effects determining the formation process of the double strands like influence of solvent, π-stacking, and backbone topology.

To include these interaction into our model we plan force field calculation which are able to describe the complete double strand including the surrounding water molecules. In a first step we will examine the accuracy of existing force fields. This is necessary since it is well known that the existing force fields are optimized for canonical bases but possess deficiancy in the description of non-canonical bases. This is already seen in the tautomeric forms of xanthine which are artificially described lower in energy than the diketo form. For an adjustment of the force field to the present problem we will use the data computed up to now. In the final step we will use the optimized force field to simulate the dissociation

process of the complete double strand including the solvent.
For our calculations we used approximately 5800 CPU-h's on the CRAY T3E-256 and T3E-512 machines.

References

1. W. Saenger. *Principles of Nucleic Acid Structure*. Springer, New York, 1984.
2. A. R. Fersht. *Trends. Biochem. Res.*, 12:301, 1987.
3. J. Šponer and P. Hobza. *J. of Mol. Structure.*, 388:115, 1996.
4. F. H. Martin, M. M. Castro, F. Aboul-ela, and I. Tinoco Jr. *Nucleic Acid Res.*, 13:8927, 1985.
5. Y. Kawase, S. Iwai, H. Inoue, K. Minura, and E. Ohtsuka. *Nucleic Acid Res.*, 14:7727, 1986.
6. D. H. Turner, N. Sugimoto, R. Kierzek, and S. D. Dreiker. *J. Am. Chem. Soc.*, 109:3783, 1987.
7. R. M. Wing, H. R. Drew, T. Takano, C. Broka, S. Tanaka, K. Itakura, and R. E. Dickerson. *Nature*, 287:755, 1980.
8. I. R. Gould and P. R. Kollmann. *J. Am. Chem. Soc.*, 116:2493, 1994.
9. Y.-P. Pang, J. L. Miller, and P. A. Kollmann. *J. Am. Chem. Soc.*, 121:1717, 1999.
10. R. A. Friedman and B. Honig. *Biophys. J.*, 69:1528, 1995.
11. L. F. Newcomb and S. H. Gellman. *J. Am. Chem. Soc.*, 116:2993, 1994.
12. D. B. Smithrud, T. B. Wyman, and F. A. Diederich. *J. Am. Chem. Soc.*, 113:4520, 1991.
13. V. M. Rotello, E. A. Viani, G. Deslongchamps, B. A. Murray, and J. Rebek. *J. Am. Chem. Soc.*, 115:797, 1993.
14. C. A. Hunter. *J. of Mol. Biol.*, 230:1025–1054, 1993.
15. P. E. Nielsen, M. Egholm, R. H. Berg, and O. Buchardt. *Science*, 254:1497, 1991.
16. P. Wittung, S. K. Kim, O. Buchardt, P. Nielsen, and B. Norden. *Nucleic Acid Res.*, 22:5371, 1994.
17. M. Egholm, O. Buchardt, L. Christensen, C. Behrens, S. M. Freier, D. A. Driver, R. H. Berg, S. K. Kim, B. Norden, and P. E. Nielsen. *Nature*, 365:566, 1993.
18. B. Hyrup and P. E. Nielsen. *Bioorg. Med. Chem.*, 4:5, 1996.
19. O. Buchardt, M. Egholm, R. H. Berg, and P. E. Nielsen. *Trends. Biotechnol.*, 11:384, 1993.
20. U. Diederichsen. *Lineare Nucleinsäure-Analoga mit peptidischem Rückgrat*. Habitilationsschrift, TU München, 1999.
21. M.F.H. Hoffmann, A.M. Brückner, T. Hupp, B. Engels, and U. Diederichsen. *Helv. Chim. Acta*, 83:2580, 2000.